

УДК 57.087

**ИНТЕГРАЛЬНАЯ ОЦЕНКА СОСТОЯНИЯ
СЛОЖНЫХ БИМЕДИЦИНСКИХ СИСТЕМ**

В.А. Фокин

Сибирский государственный
медицинский университет, г. Томск
E-mail: fokin@ssmu.ru

Фокин Василий Александрович, д-р техн. наук, профессор кафедры медицинской и биологической кибернетики Сибирского государственного медицинского университета, г. Томск.
Email: fokin@ssmu.ru
Область научных интересов: многомерный анализ биомедицинских данных, математическое моделирование в медицине и биологии, оценка состояния биосистем.

Предложена информационная технология интегральной оценки состояния биосистем, обеспечивающая единство оценочного алгоритма на различных уровнях их структурно-функциональной организации.

Ключевые слова:

Информационные технологии, интегральная оценка состояния биосистем, информативность показателей, статистическое моделирование.

Key words:

Information technology, integral estimation of bio-systems state, self-descriptiveness characteristics, statistical modeling.

Введение

Решение проблемы оценки состояния сложных биосистем существенным образом зависит от возможностей количественного описания протекающих в них процессов в рамках ограниченной априорной информации в условиях многообразия внешних воздействий и индивидуальных особенностей исследуемого биообъекта. С этой точки зрения, важная роль при разработке информационных технологий оценки состояния биосистем отводится формированию, накоплению и использованию массивов многомерных медико-биологических данных [1–3].

В целом, в медико-биологических исследованиях складывается достаточно противоречивая ситуация. С одной стороны, накоплены разнообразные массивы данных, отражающие всевозможные встречающиеся в медицинской науке и клинической практике ситуации, а с другой – непропорционально малое количество информации, получаемое из их анализа. Это связано с тем, что, несмотря на очевидные успехи, использование математических методов и вычислительной техники в ряде случаев оказывается недостаточно эффективным с точки зрения прикладных целей исследования: попытки точного описания приводят к чрезвычайно сложным для анализа математическим моделям, а данные, полученные в отдельных исследованиях не позволяют получать надежные оценки состояния в условиях широкой внутри- и межиндивидуальной вариабельности биообъектов, характеризующих исследуемое состояние. Как результат, при хороших теоретических построениях практическое применение математических моделей и алгоритмов для количественной оценки состояния биосистем может приводить к широкому разбросу в величине и надежности получаемых оценок.

Существенной особенностью большинства измеряемых медико-биологических показателей является то, что при тех или иных изменениях состояния биосистемы они носят разнонаправленный характер, при этом ряд из них находятся в пределах или незначительно выходят за границы статистической нормы. Это создает определенные трудности в интерпретации и оценке состояния системы, поскольку при изменении условий

функционирования некоторые показатели становятся «лучше», другие – «хуже». Оценить состояние системы, выделяя какие-либо конкретные из них, бывает достаточно трудно, поэтому особую важность приобретают обобщенные критерии оценки состояния. Более того, такие интегральные оценки состояния, построенные по совокупности многомерных данных, оказываются более чувствительны даже к малым изменениям состояния биосистем [4–6].

В работе рассматривается подход к интегральной оценке состояния сложных биосистем на основе концептуальной модели системы многомерных данных.

Концептуальная модель системы биомедицинских данных

Результаты медико-биологических экспериментов и клинических исследований всегда имеют естественную концептуальную основу для систематизации и интеграции: они относятся к одному и тому же объекту – человеческому организму, поэтому любые наблюдения или измерения не теряют своей значимости и всегда несут в себе потенциальную возможность получения новой информации. Это предъявляет определенные требования к организации систем накопления данных и использованию математических методов их совместного анализа. В модели системы данных должны учитываться возможности по описанию не только непосредственно накопленных данных, но и возможности, необходимые для выявления обобщенных свойств биосистемы.

Проведение любого исследования начинается с выдвижения научной гипотезы (постановки цели исследования), для адекватного обоснования и проверки которой выявляются параметры биосистемы, подлежащие количественному или качественному измерению, а также инструментальные и аналитические методы, позволяющие оценивать и анализировать их с заданной точностью. Отсюда следует, что задание цели исследования полностью определяет систему сбора данных и способы их совместного анализа с тем, чтобы обеспечить необходимую степень достоверности и полноценности получаемых результатов.

Определим концептуальную модель M интегрированной системы данных (ИСД) [7]:

$$M = \{P, G, R, Q, C\},$$

где $P = \{P_i | i \in N_p\}$ – множество требований методологического характера; $G = \{G_i | i \in N_g\}$ – множество групп пользователей; $R = \{R_i | i \in N_r\}$ – множество целей пользователей; $Q = \{Q_i | i \in N_q\}$ – множество семантических моделей предметной области; $C = \{C_i | i \in N_c\}$ – множество каналов идентификации (наблюдения) объекта исследования; N_p, N_g, N_r, N_q, N_c – количество элементов соответствующих множеств.

Структурная схема концептуальной модели представлена на рис. 1. Система данных приобретает обобщенные (интегральные) свойства в том случае, если она формируется при условии учета и соответствия данных свойствам элементов перечисленных выше множеств. К методологическим требованиям P_i формирования интегрированной системы данных будем относить выполнение условий:

- отделения данных от содержащейся в них информации и отделения процессов обработки данных от процессов обработки информации. Требование независимости системы данных от моделей их анализа и интерпретации обеспечивает возможность многократного использования данных на различных уровнях формализованного логико-математического описания в исследовании биосистем и их количественном анализе.
- познавательной полезности любых экспериментальных данных в описании состояния системы. Модель системы ближе к действительности, если изучается много ее особенностей, однако содержание тех или иных данных и степень их вклада в построение и интерпретацию модели системы существенным образом зависит от выдвигаемой цели исследования.

Пользователи интегрированной системы данных G_i представлены двумя основными подгруппами: пользователи, пополняющие систему данными, и пользователи, использующие

накопленные в системе данные. Различие между ними заключается в направлении потока данных: цель пользователей первой подгруппы – наполнение ИСД результатами экспериментов (поток данных направлен от изучаемой биосистемы в систему данных), а второй – использование накопленных данных для управления и прогнозирования поведения объекта исследования (поток направлен от системы данных к изучаемой биосистеме).

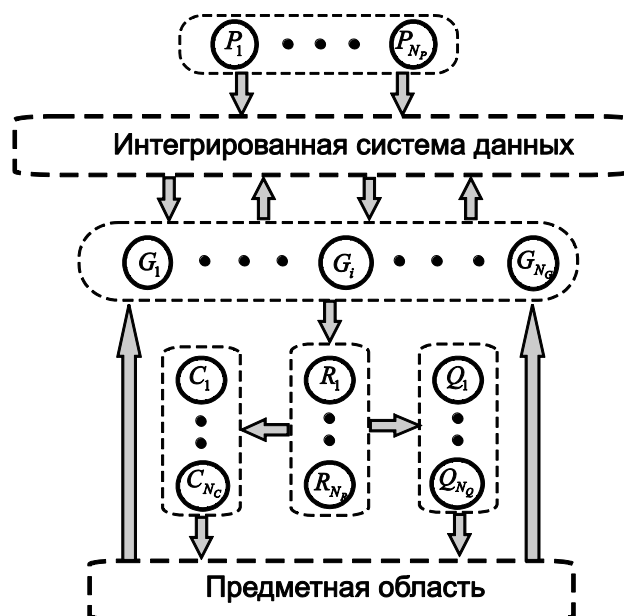


Рис. 1. Схема концептуальной модели интегрированной системы данных

Область исследования может быть представлена различными семантическими моделями, адекватными потребностям пользователей системы данных, реализующим свои определенные цели R_i . Не конкретизируя предметную область исследований, в самом общем виде можно выделить три вида целей пользователей:

- оценка состояния биосистемы, на изучение которой направлена деятельность;
- выявление причинно-следственных связей (функционирования) биосистемы;
- выбор оптимального воздействия на биосистему.

Семантическая модель Q_i находит свое выражение в установлении смыслового соответствия двух множеств: множества переменных $\vec{x} = \{x_i | i \in \overline{1, m_x}\}$ – свойств, определяющих исследуемую систему S в терминах некоторой семантической модели и множества параметров $\vec{b} = \{b_i | i \in \overline{1, m_b}\}$ – свойств, характеризующих различия в проявлениях переменных x_i , здесь m_x и m_b – количество элементов соответствующих множеств. Каждая переменная x_i имеет свое уникальное имя, определяемое семантической моделью предметной области, и представлена множеством количественных или качественных величин X_i^B – данных, однозначно определяемых декартовым произведением параметров $B = b_1 \times b_2 \times \dots \times b_{m_b}$. Таким образом, наблюдение над объектом исследования представляет собой упорядоченную пару, единственным образом отображающую значение параметрического множества во множество значений переменных, т. е. в самом общем виде данные представляют собой функцию:

$$f: B \rightarrow X^B, \text{ где } X^B = \bigcup_i X_i^B.$$

Выбор параметров имеет существенное значение для эффективного описания свойств системы [8]. В качестве параметров системы наиболее часто используются время,

пространство, группа или их комбинации. В целом параметры системы задаются исходя из следующих общих требований. Параметры должны:

- соответствовать свойствам (переменным) исследуемой системы;
- соответствовать целям изучения системы;
- однозначно определять значение свойства системы.

Данные, полученные в ходе отдельного эксперимента можно определить как некоторое подмножество $X_i^{C,R}$ множества X_i^B , которое определяется доступными исследователю каналами наблюдения C и заданными целями исследования R и может быть записано в виде пары:

$$X_i^{C,R} = ((\vec{x}_i^*, X^C), (\vec{b}_i^*, B^R)),$$

где \vec{x}_i^* и \vec{b}_i^* – векторы, задающие некоторые подмножества переменных и параметров, определяемых семантической моделью Q_i ; X^C и B^R – множества значений переменных и параметров определяемых процедурой измерения через конкретный канал наблюдения C в соответствии с целью исследования R .

Определенное таким образом множество $X_i^{C,R}$ определяет множество данных полученных в исследовании. Учитывая, что значения переменных множества \vec{x}_i могут быть получены через различные каналы наблюдения C_i и на разных параметрических множествах \vec{b}_j , задаваемых целями пользователей R_j , полную систему данных X^Q соответствующих семантической модели Q можно определить следующим образом:

$$X^Q = \bigcup_{C,R} X_i^{C,R}.$$

Появление новых и совершенствование имеющихся методов медико-биологических и клинических исследований приводит к постоянному качественному и количественному увеличению объема клинико-лабораторных данных, характеризующих оценку состояния биомедицинских систем. В этих условиях появляется возможность постановки и решения вновь возникающих исследовательских задач на уже накопленных массивах данных. Принципиальная возможность такого подхода обусловлена тем, что цель любого нового исследования R формализуется через задание набора подлежащих измерению переменных \vec{x} на параметрическом множестве \vec{b} в соответствии с определенной семантической моделью Q объекта исследования и выбранными измерительными каналами C . Если имеющаяся в распоряжении исследователя система данных X^Q позволяет синтезировать соответствующий набор значений $X^{C,R} \subseteq X^Q$, то он будет являться адекватным отображением объекта исследования и может быть использован для изучения его свойств.

В определенной подобным образом системе данных наиболее полным образом отображается информация, полученная об объекте исследования в результате всей совокупности экспериментальных исследований, наблюдений и т. п.

Интегральная оценка состояния

Методы получения количественной оценки состояния сложных биосистем определяются выбором набора показателей, используемых для ее расчета. На практике наиболее часто встречаются следующие подходы:

- Выбор в качестве оценки состояния небольшого набора независимых показателей, в идеальном случае, одного. При этом количественной оценкой состояния системы служит величина самого непосредственно измеряемого показателя. Однако для сложных систем

обнаружить такие показатели, которые бы однозначно определяли ее состояние, бывает достаточно трудно.

- Формирование оценок состояния с использованием методов многомерного статистического анализа данных. Сюда можно отнести всевозможные процедуры многомерного регрессионного, дискриминантного, факторного и других многомерных методов анализа. Достаточно эффективно для получения оценок состояния последовательное применение нескольких статистических процедур анализа [9, 10]. При этом в качестве оценок состояния используются характеристики, следующие из соответствующего метода статистического анализа.

Использование интегральных показателей, формируемых на основе анализа закономерностей функционирования изучаемых биосистем. Это подходы, связанные с построением полуэмпирических индексов состояния, таких как биохимический, клинический индексы; получением оценок функционального состояния организма (уровень функционирования, функциональный резерв, степень напряженности регуляторных механизмов); анализом вербальных данных на основе теории нечетких множеств, применением методов многомерного шкалирования, нейросетевых технологий и т.п. [11–13]. Их можно рассматривать как модельно-независимые оценки, характеризующие систему в целом.

Задача нахождения интегральной оценки сводится к выбору признаков, характеризующих состояние биосистемы и построению функционального отображения их значений в одномерную шкалу. Следует отметить, что оценка состояния S биосистемы всегда производится по отношению к ее эталону состояния S_0 . В качестве последнего для человека может быть выбрано, например, состояние здорового организма.

Определение набора показателей x_i , по которым должна производиться такая оценка – трудно формализуемая задача. В целом можно сформулировать следующие общеметодологические требования к построению критериев интегральной оценки состояния биомедицинских систем:

- Математический аппарат, используемый для построения интегрального критерия, должен учитывать вариабельность и наличие взаимосвязей между анализируемыми показателями.
- В интегральный критерий состояния системы должен включаться весь комплекс показателей, характеризующих определенный уровень в иерархическом описании системы.
- Функциональный вид интегрального критерия должен обеспечивать единство оценочного алгоритма для различных целевых состояний системы.
- Интегральный критерий должен позволять проводить индивидуализированную оценку состояния объекта исследования.

Пусть референтное (эталонное) состояние биосистемы S_0 представлено множеством N_{S_0} объектов, у которых проведено измерение показателей $\{\bar{x}_i | i \in N_{S_0}\}$. Количественно величина интегральной оценки близости объекта $\bar{y}_j \in S$ к эталону состояния может быть определена следующим образом [14]:

$$I_{S_0}(\bar{y}_j) = \frac{d(\bar{y}_j, S_0)}{D_{S_0}}, \quad (1)$$

где $d(\bar{y}_j, S_0)$ – мера близости объекта \bar{y}_j к эталонному состоянию S_0 ; D_{S_0} – мера компактности области, занимаемой в пространстве признаков объектами, относящимися к эталонному состоянию.

В качестве меры близости объекта \bar{y}_j к эталонному состоянию S_0 использовалось усредненное расстояние Махаланобиса от этого объекта до всех объектов состояния S_0 :

$$d(\bar{y}_j, S_0) = \frac{1}{N_{S_0}} \sum_{i=1}^{N_{S_0}} d(\bar{y}_j, \bar{x}_i).$$

Параметр D_{S_0} зададим в виде внутримножественного расстояния [15]:

$$D_{S_0} = \frac{1}{N_{S_0}} \sum_{i=1}^{N_{S_0}} \frac{1}{N_{S_0} - 1} \sum_{l=1}^{N_{S_0} - 1} d(\vec{x}_i, \vec{x}_l),$$

т. е., как усредненное значение средних расстояний от каждого объекта, относящегося к состоянию S_0 , до всех оставшихся. Нормировка на величину D_{S_0} в выражении (1), позволяет учесть как конфигурацию области S_0 , так и взаимное расположение объектов в ней.

Основная проблема при использовании критериев, основанных на многомерных методах анализа данных, обусловлена малыми объемами выборок, характеризующих эталонное состояние, что приводит к значительной вариабельности оценок, получаемых с их использованием. Это является отражением специфики биомедицинских данных и прежде всего их широкой внутри- и междуиндивидуальной вариабельности. Вследствие чего проведение повторных измерений на даже одной и той же выборке может приводить к различным количественным значениям оцениваемых характеристик. Оценка статистических свойств предложенного выше интегрального критерия (1) представляет собой нетривиальную задачу, решение которой с использованием только аналитических подходов обусловлено значительными трудностями, а ряде практических случаев невозможно.

В этом случае исследование статистических свойств интегрального критерия может быть эффективно реализовано методами статистического моделирования, результаты которого позволяют, с одной стороны, численно оценивать статистические свойства критерия, а с другой – позволяют определить условия, накладываемые на условия формирования референтных выборок, необходимые для получения устойчивых оценок.

Оценка статистических характеристик интегрального показателя проводилась в два этапа. На первом этапе моделировалось M выборочных множеств X_k ($k = \overline{1, M}$) заданного объема, соответствующих статистическим характеристикам референтного состояния S_0 . Полученные последовательности значений имитируют взятие выборок из одной и той же совокупности и, следовательно, будут свободны от погрешностей, обусловленных влиянием внутри- и междуиндивидуальной вариабельности биологических данных. Далее для каждого множества X_k вычислялись величины оценок $I_{S_0,k}(\vec{x})$, распределение которых в дальнейшем использовалось для исследования статистических свойств интегрального критерия. Здесь вектор \vec{x} характеризует объект, для которого производится оценка.

В зависимости от того, известен или нет закон распределения многомерных данных, для статистического моделирования наборов их значений могут применяться различные методы [16, 17]. Статистические свойства интегрального критерия оценивались путем вычисления среднего значения интегрального показателя

$$\hat{I}_{S_0}(\vec{x}) = \frac{1}{M} \sum_{k=1}^M I_{S_0,k}(\vec{x}),$$

среднего квадратичного отклонения

$$\hat{\sigma}_I(\vec{x}) = \frac{1}{M} \sum_{k=1}^M \left(I_{S_0,k}(\vec{x}) - \hat{I}_{S_0}(\vec{x}) \right)^2.$$

Для оценки вариабельности интегрального показателя рассчитывался коэффициент вариации

$$V = \frac{\hat{\sigma}_I(\vec{x})}{\hat{I}_{S_0}(\vec{x})} \cdot 100 \, \%.$$

и $(1-p) \cdot 100 \, \%$ -го доверительный интервал, как интервал, содержащий значения I_{S_0} находящиеся между $p/2 \cdot 100 \, \%$ и $(1-p/2) \cdot 100 \, \%$ числом всех значений интегрального показателя в ранжированном ряду оценок. Здесь p – соответствующий уровень статистической значимости. Такой непараметрический способ оценки доверительного интервала позволяет

оценивать его без каких-либо предположений относительно вида закона распределения и статистических свойств интегрального показателя.

Информационная технология интегральной оценки

Компьютерные технологии и разрабатываемые на их основе программно-технические средства и информационные системы являются универсальным инструментом, используемым в биомедицинских исследованиях и практическом здравоохранении для формирования структурированных массивов данных, их анализа и принятия решений на основе алгоритмов обработки, в соответствии с целями пользователя [1, 2]. Эффективность решения задач интегральной оценки состояния биосистем во многом будет зависеть от разработки соответствующих программно-технических средств, и возможностей их применения к массивам данных в конкретных областях медико-биологических исследований. Опишем структуру и компьютерную технологию интегральной оценки состояния биосистем, в основу которой положена описанная выше модель интегрированной системы данных [18, 19].

В соответствии с предлагаемой методологией оценки состояния биосистем, программная и диалоговая архитектура компьютерной системы включает в себя блоки по реализации следующих задач:

- формализованное описание измеряемых показателей предметной области исследований;
- наполнение базы данных результатами исследований;
- интегральная оценка состояния биосистем на основе анализа сформированных по совокупности результатов различных исследований референтных выборок; Реализация данных задач, с точки зрения разработки средств накопления и переработки данных, является далеко не тривиальной проблемой.

Специфика их решения обусловлена тем, что сами данные [2]:

- имеют неограниченный объем;
- являются разнородными (количественными, качественными, текстовыми);
- результаты анализа данных должны быть понятны и конкретны;
- инструменты для обработки «сырых» данных должны быть просты в использовании.

Назначением разрабатываемого инструментального средства является формирование банка медико-биологических данных с целью оптимизации и поддержки научно-исследовательской деятельности по комплексному анализу и оценке состояния сложных биосистем. Отличительной особенностью предлагаемого подхода является то, что накопление массивов разнородных показателей в базе данных осуществляется с максимальным учетом атрибутов, характеризующих изучаемое состояние. Это позволяет в дальнейшем эффективно использовать имеющуюся информацию для агрегации референтных массивов данных для получения интегральных оценок состояния исследуемых биосистем.

Разработанная система (рис. 2) ориентирована на выполнение следующих функций:

- ввод и хранение экспериментальных данных в базе данных;
- генерацию бланков ввода данных;
- формирование выборок референтных данных из базы данных;
- статистическое моделирование массивов данных референтных выборок;
- вычисление интегральных оценок состояния.

Описание предметной области исследований содержится в базе описаний показателей, которая дополняется пользователем по мере необходимости. Единицей записи в ней является описание измеряемого показателя или условий проведения измерения. В базе данных пациентов содержатся такие атрибуты пациента, которые не изменяются от исследования к исследованию или изменяются крайне редко.

База описаний показателей является основой для создания бланков исследований и запросов по формированию референтных выборок. Для этого в разработанной системе предусмотрены модули формирования бланков исследований и запросов. Их включение в систему позволяет наиболее гибко учитывать все разнообразие цели проводимых исследований.

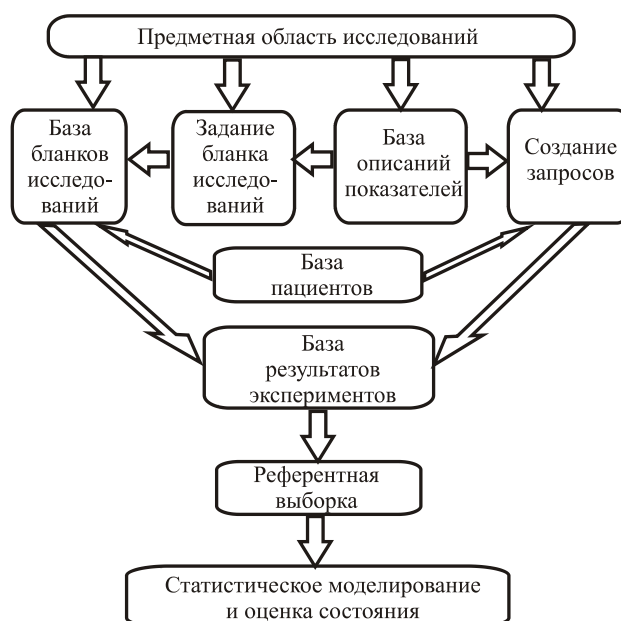


Рис. 2. Структурная схема программы

Модуль формирования бланков исследований позволяет создавать формы бланков для занесения результатов реального эксперимента, наиболее последовательно и полно отображающих содержание проводимого исследования. Для каждого бланка создается отдельная таблица в базе результатов экспериментов, в которой хранятся все результаты измерений. Модуль формирования запросов ориентирован на проведение «виртуального» исследования, т. е. задание показателей и атрибутов состояния пациентов, для которых референтная выборка данных будет сформирована из имеющейся базы результатов исследований. Модуль статистического моделирования и оценки состояния предназначен для получения интегральных оценок состояния биосистемы.

Кроме перечисленных выше функций в компьютерной системе предусмотрен также ряд дополнительных функций, обеспечивающих удобство взаимодействия с пользователем: экспорт данных в основных форматах, ведение и сохранение протокола результатов анализа, встроенная помощь и т. п.

Примеры реализации

Приведем результаты интегральной оценки состояния некоторых биосистем на основе предложенной технологии.

Оценка системы красной крови при различных патологиях. Системные свойства красной крови проявляются в том, что ее структурно-функциональные изменения возникают не только в результате нарушений функций отдельных ее компонентов, но и при нарушениях функций других органов и систем организма, что делает систему красной крови одним из основных индикаторов происходящих в организме изменений. Исходными данными для проведения статистических оценок послужили данные сканирующей электронной микроскопии (СЭМ) поверхностной архитектуры клеток красной крови, полученные коллективом авторов под руководством академика РАМН В.В. Новицкого [20, 21] по результатам обследований онкологических больных, а также здоровых лиц. Поскольку форма эритроцитов и их способность к деформации является следствием комплекса нарушений свойств, организации и метаболизма отдельных компонентов эритроцитов, обусловленных наличием соответствующего патологического процесса, то данные СЭМ могут быть использованы для интегральной оценки степени изменений происходящих в системе красной крови по выбранному комплексу показателей.

На рис. 3, а, представлены результаты расчетов при различных локализациях одного и того же патологического процесса. Эталонное состояние – выборка здоровых людей, оцениваемые

состояния – пациенты с первичным раком III-IV стадий различных локализаций: легкие (РЛ), область головы и шеи (РОГШ), желудок (РЖ), тонкая кишка (РТК). Видно, что в зависимости от локализации опухолевого процесса степень изменений со стороны красной крови существенно различается. В частности, при раке органов пищеварительной системы степень изменений в системе красной крови значительно выше, чем при раке легкого и опухолях головы и шеи. В тоже время видно (рис. 3, б), что величина оценки изменений в системе красной крови при третьей и четвертой степени опухолевого процесса меняется гораздо меньше.

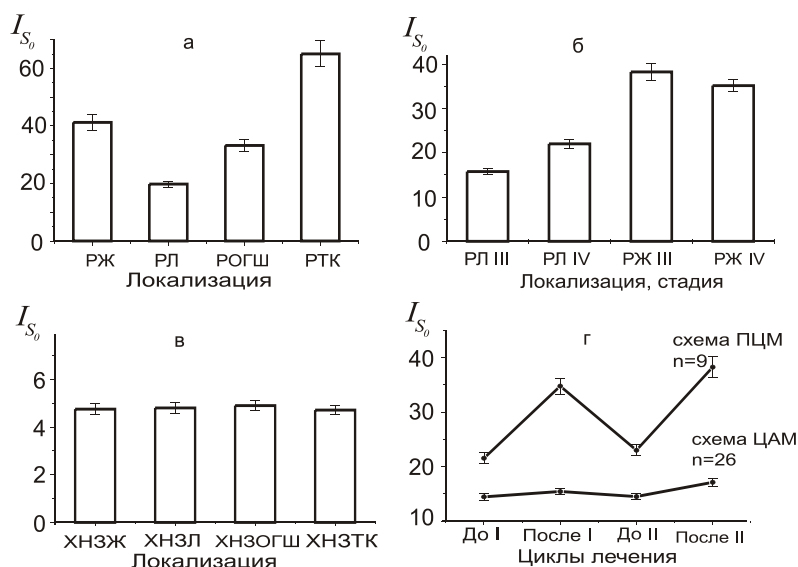


Рис. 3. Интегральная оценка изменений системы красной крови при различных патологических состояниях (параметры моделирования: $N_{S_0} = 1000$, $M = 400$)

На рис. 3, в, представлена оценка степени изменений системы красной крови при неспецифических хронических заболеваниях с локализацией, аналогичной рассмотренным выше опухолевым процессам: ХНЗЖ – хронические неспецифические заболевания желудка, ХНЗЛ – хронические неспецифические заболевания легких, ХНЗОГШ – хронические неспецифические заболевания области головы и шеи, ХНЗТК – хронические неспецифические заболевания тонкой кишки. Несмотря на то, что как и в случае а, затрагиваются одни и те же подсистемы организма, однако количественно величина интегральной оценки выявленных изменений в системе красной крови при хронических неспецифических заболеваниях оказалась существенно ниже, чем при опухолях соответствующих локализаций.

Интегральная оценка изменений системы красной крови в динамике лечения рака легких химиотерапией при одной из двух стандартных схем назначения цитостатических препаратов: ЦАМ (циклофосфан-адриамицин-метатрексат) или ПЦМ (проспидин-циклофосфат-метатрексат) представлена на рис. 3, г. Анализ кривых показывает, что при применении схемы ПЦМ состояние системы крови в целом ухудшается по окончании каждого из циклов, в отличие от схемы ЦАМ, при которой существенных сдвигов в динамике полученной оценки не отмечается, что можно интерпретировать как более щадящее воздействие схемы ЦАМ на систему красной крови в целом. Однако нельзя исключать, что в первую группу могли включаться пациенты с исходно более выраженными изменениями в системе красной крови и, соответственно, более низким уровнем адаптационных возможностей.

В целом, полученные на основе статистического моделирования данных интегральные оценки состояния системы красной крови хорошо согласуются с данными клинических исследований о тяжести анализируемых состояний.

Оценка изменений спектра поглощения газовыделений биосистем. В настоящее время методы спектрального анализа состава газовыделений биосистем приобретают все более важное значение, как неинвазивные методы, способные выявлять изменения в биосистемах на

ранних стадиях происходящих в них изменений [22, 23]. Газовыделения, являющиеся результатом жизнедеятельности биообъектов, имеют сложный многокомпонентный состав, что приводит к существенным затруднениям в интерпретации экспериментальных спектроскопических данных традиционными методами решения обратной спектроскопической задачи и, как следствие, неопределенности последующей оценки состояния биообъекта.

При анализе состава газовой смеси, использование CO_2 -лазера в качестве источника излучения, позволяет реализовать до 100 линий генерации в диапазоне 910-980 и 1040-1090 см^{-1} . Таким образом, за один измерительный цикл можно получить распределение (скан) коэффициента поглощения исследуемой смеси газов на дискретном наборе длин волн в области 9...11 мкм. К достоинствам такого метода регистрации спектра относятся высокое спектральное разрешение, высокая чувствительность по показателю поглощения, отсутствие сигнала, если спектральная линия излучения находится вне линий поглощения исследуемой газовой смеси и малый объем требуемого образца газа [24].

Вектор состояния \vec{x} задавался интенсивностями поглощения на линиях генерации CO_2 -лазера в диапазонах 932...954 и 966...982 см^{-1} . Выбор данных диапазонов длин волн был обусловлен тем, что в нем генерация характеристик сигнала была наиболее устойчива и воспроизводима. В каждом из диапазонов были измерены интенсивности поглощения на 12 линиях. Состояние биосистемы рассматривалось как точка в двумерном пространстве (I_1, I_2) интегральных оценок, соответствующем указанным диапазонам длин волн [25].

На рис. 4, а, приведены результаты интегральной оценки состава выдыхаемого воздуха при некоторых бронхо-легочных заболеваниях: БА – бронхиальная астма, ХОБЛ – хроническая обструктивная болезнь легких, ПНЕВМ – пневмония и ИБС – ишемическая болезнь сердца.

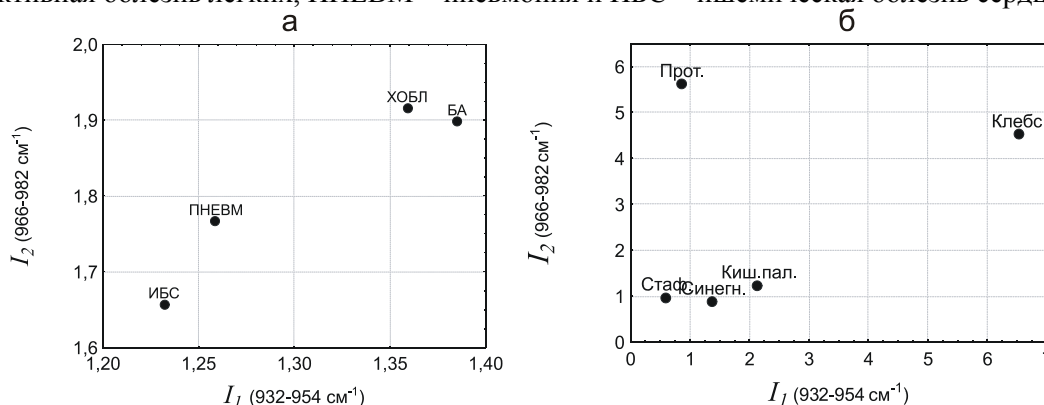


Рис. 4. Интегральная оценка изменений состава газовыделений (параметры моделирования: $N_{S_0} = 1000$, $M = 400$)

Референтное состояние составили пробы выдыхаемого воздуха здоровых лиц. Полученные интегральные оценки могут быть интерпретированы с медицинской точки зрения [26]. Бронхиальная астма и ХОБЛ относятся к группе хронических болезней дыхательных путей со сходными патологическими механизмами и близкими по клинике симптомами. Видно, что точки, соответствующие этим состояниям, расположены близко друг к другу, причем значительно дальше от референтного состояния, чем при пневмонии, что можно интерпретировать, проявление системных сдвигов в при переходе воспаления в хроническую фазу. При ишемической болезни сердца изменения в бронхо-легочной системе происходят в значительно меньшей степени, что и нашло отражение в меньшем отклонении интегральной оценки состава выдыхаемого воздуха от выдыхаемого воздуха группы здоровых людей по сравнению с другими анализируемыми состояниями.

Особенности метаболизма различных культур бактерий обусловлены характерным для каждой из них набором ферментов, которые, в свою очередь, определяют специфику биохимических процессов, сопровождающих развитие колонии. В процессе происходящего метаболизма происходит выделение различных газовых компонент, обнаружение которых (как

качественное, так и количественное) является актуальной задачей с точки зрения контроля загрязнений различных объектов болезнетворными бактериями. На рис. 4, б, приведены результаты оценки степени изменений состава газовыделений бактерий помещенных в питательный бульон (Клебс – клебсиелла, Прот – протей, Стаф – стафилококк, Киш.пал. – кишечная палочка, Синегн. – синегнойная палочка). Референтное состояние: газовыделения чистого бульона. Видно, что точки, соответствующие разным типам бактерий располагаются в различных областях пространства интегральных оценок. В частности, интегральная оценка спектра поглощения газовыделений бактерий протей и клебсиеллы значительно отличается от оценок, полученных для бактерий синегнойной и кишечной палочек и стафилококка, что может иметь определенное диагностическое значение. Поскольку регистрация газовыделений бактерий возможна на самых ранних стадиях развития колоний, то это особенно важно для разработки методов ранней диагностики болезней с высоким риском заражения окружающих, таких как туберкулез, легионеллез, микоплазмоз и др., а также для предупреждения эпидемических вспышек.

Выводы

1. На основе системного анализа свойств биомедицинских данных предложена концептуальная модель интегрированной системы данных, при разработке которой учитывались системные требования, предъявляемые к формированию массивов биомедицинских данных, существенные для оценки свойств биосистем.
2. Предложен критерий и алгоритмы интегральной оценки состояния биосистем, позволяющие проводить оценку степени происходящих в ней изменений по отношению к заданному референтному состоянию, по совокупности многомерных данных.
3. Разработан комплекс проблемно-ориентированных компьютерных программ для статистического моделирования и интегральной оценки состояния биосистем. Программная и диалоговая архитектура включает в себя программные блоки по реализации следующих задач: блок формализованного описания измеряемых показателей предметной области исследований, блок наполнения базы данных результатами экспериментов и блок интегральной оценки состояния на основе анализа референтных выборок, сформированных по совокупности результатов различных экспериментов.

Работа выполнена при частичной финансовой ФЦП (г/к. № 07.514.11.4054).

СПИСОК ЛИТЕРАТУРЫ

1. Генкин А.А. Новая информационная технология анализа медицинских данных (программный комплекс ОМИС). – СПб.: Политехника, 1999. – 191 с.
2. Дюк В., Эммануэль В. Информационные технологии в медико-биологических исследованиях. – СПб.: Питер, 2003. – 528 с.
3. Богомолов А.В., Гридин Л.А., Кукушкин Ю.А., Ушаков И.Б. Диагностика состояния человека: математические подходы. – М.: Медицина, 2003. – 464 с.
4. Конрадов А.А. Статистические подходы к анализу многомерных гетерогенных биологических систем // Радиационная биология, радиоэкология. – 1994. – Т. 34. – Вып. 6. – С. 877–886.
5. Попова Е.В., Струнникова Е.В. Применение методов многомерного статистического анализа в гигиенических исследованиях // Гигиена и санитария. – 1992. – № 3. – С. 63–64.
6. Боженко В.К., Беридзе А.Д., Шишкин А.М., Гуслистый В.П. Применение многопараметрических методов для анализа лабораторных показателей крови // Клин. лаб. диагностика. – 1997. – № 10. – С. 45–47.
7. Фокин В.А. Концептуальная модель системы биомедицинских данных // Вестник новых медицинских технологий. – 2005. – Т. 12. – № 1. – С. 118–120.
8. Клар Дж. Системология. Автоматизация решения системных задач / Пер. с англ. – М.: Радио и связь, 1990. – 554 с.

9. Миронкина Ю.Н., Бобров А.Ф. Информационная технология статистического синтеза критериев и алгоритмов оценки функционального состояния человека в прикладных медико-биологических исследованиях // Информационные технологии. – 1998. – № 3. – С.41–47.
10. Подвальный С.Л., Матасов А.С., Бырко И.А. Методы многомерной классификации в задачах медицинской диагностики // Машиностроитель. – 2002. – № 8. – С. 59–61.
11. Нисевич Н.И., Марчук Г.И., Зубикова И.И., Погожев И.Б. Математическое моделирование вирусного гепатита. – М.: Наука, 1981. – 352 с.
12. Баевский Р.М. Оценка и классификация уровней здоровья с точки зрения теории адаптации // Вестник АМН СССР. – 1989. – № 8. – С. 73–78.
13. Муха Ю.П., Скворцов М.Г., Авдеюк О.А. и др. Диагностический комплекс основных жизненно важных функций человека по интегральному параметру на основе нейросетевых технологий // Биомедицинские технологии и радиоэлектроника. – 2001. – № 4. – С. 38–41.
14. Фокин В.А. Критерий оценки состояния сложных биосистем // Известия Томского политехнического университета. – 2004. – Т. 307. – № 5. – С. 136–138.
15. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 416 с.
16. Ермаков С.М., Михайлов Г.А. Статистическое моделирование. 2-е изд. – М.: Наука, 1982. – 296 с.
17. Bryan F. J. Randomization, Bootstrap and Monte Carlo Methods in Biology – London: Chapman and Hall/CRC, 1997. – 424 p.
18. Свид. об официальной регистрации. Программа для ЭВМ «StatSys» / В.А. Фокин И.С. Хакимов О.Ю. Никифорова. № 2006614010; заявл. 29.09.2006; зарегистр. 2.11.2006. – М.: РОСПАТЕНТ, 2006.
19. Свид. об официальной регистрации. Программа для ЭВМ «ИСД Биомед» / В.А. Фокин, А.В. Дубровин. – № 2007612525; заявл. 27.04.2007; зарегистр. 15.06.2007. – М.: РОСПАТЕНТ, 2007.
20. Новицкий В.В., Степовая Е.А., Гольдберг В.Е. и др. Обратимая агрегация и поверхностная архитектура эритроцитов периферической крови у больных раком легкого до и в ходе проведения противоопухолевой полихимиотерапии // Экспериментальная и клиническая фармакология. – 1999. – Т. 62. – № 5. – С. 28–30.
21. Новицкий В.В., Степовая Е.А., Гольдберг В.Е., Колосова М.В., Рязанцева Н.В., Корчин В.И. Эритроциты и злокачественные образования. – Томск: STT, 2000. – 288 с.
22. Филипп М.Р. Анализ выдыхаемого воздуха в медицине // В мире науки. – 1992. – № 9–10. – С. 42–48.
23. Степанов Е.В. Методы высокочувствительного газового анализа молекул-биомаркеров в исследованиях выдыхаемого воздуха // Труды ИОФАН. – 2005. – Т. 61. – С. 5–47.
24. Козинцев В.И., Белов М.Л., Городничев В.А., Федотов Ю.В. Лазерный оптико-акустический анализ многокомпонентных газовых смесей – М.: Изд-во МГТУ им. Н.Э. Баумана, 2003. – 352 с.
25. Фокин В.А., Кистенев Ю.В., Красноженов Е.П., Пономарев Ю.Н. Интегральная оценка состояния биообъектов на основе анализа спектров поглощения их газовыделений // Развитие научно-технического сотрудничества российских научных и научно-образовательных центров с учеными-соотечественниками, работающими за рубежом: Труды Междунар. научно-практ. конф. – 2–4 апреля 2010, Томск. – Томск, 2010. Электронный ресурс, 2010. URL: <http://dialog.extech.ru/tomsk/itog.php> (дата обращения 25.11.2011).
26. Пономарев Ю.Н., Кистенев Ю.В., Огородова Л.М. и др. Диагностика воспалительных процессов в органах дыхания на основе анализа состава выдыхаемого воздуха. – Томск: Изд-во ТПУ, 2007. – 181 с.

Поступила 28.11.2011 г.